

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 9, 2002, 00:09:27 ; Search time 8498.8 seconds  
(without alignments)  
30.345 Million cell updates/sec

Title: US-09-851-670-16  
Perfect score: 24  
Sequence: 1 gtccaagcagcagcattctgca 24

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues  
Total number of hits satisfying chosen parameters: 111874

Minimum DB seq length: 0  
Maximum DB seq length: 60

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_estfun:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estom:\*  
5: em\_estopl:\*  
6: em\_estba:\*  
7: em\_estro:\*  
8: em\_estov:\*  
9: em\_hc:\*  
10: gb\_est1:\*  
11: gb\_est2:\*  
12: gb\_hc:\*  
13: gb\_gss:\*  
14: em\_gss\_fun:\*  
15: em\_gss\_hum:\*  
16: em\_gss\_inv:\*  
17: em\_gss\_pln:\*  
18: em\_gss\_pro:\*  
19: em\_gss\_rnd:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	15.8	65.8	51	13	A2793014 2M0045E17
2	15.2	63.3	59	13	A2361084 1M0104D08
3	14.4	60.0	54	13	A2921737 1006031H1
4	14	58.3	39	11	BF339449 602039103
5	13.8	57.5	39	11	BF342092 602012848
6	13.8	57.5	29	13	A2658555 1M0535113
7	13.8	57.5	39	10	A2658555 1M0535113
8	13.4	55.8	43	13	A2869134 2M0181E10
9	13.4	55.8	51	10	AU013478 AU013478
10	13.2	55.0	29	13	A2588457 1M0396L17
11	13.2	55.0	42	13	A2632796 1M0487E12
12	13.2	55.0	57	10	AL595919 AL595919

Result No.	Score	Query Match	Length DB	ID	Description
13	54.2	50	10	10	AU105510 AU105510
14	54.2	50	10	10	AU105518 AU105518
15	54.2	50	10	10	AU105525 AU105525
16	54.2	50	10	10	AU105526 AU105526
17	54.2	50	10	10	AU105530 AU105530
18	54.2	50	10	10	AU105531 AU105531
19	54.2	50	10	10	AU105533 AU105533
20	54.2	50	10	10	AU105534 AU105534
21	54.2	50	10	10	AU105535 AU105535
22	54.2	50	10	10	AU105536 AU105536
23	54.2	50	10	10	AU105537 AU105537
24	54.2	50	10	10	AU105539 AU105539
25	54.2	50	10	10	AU105540 AU105540
26	54.2	50	10	10	AU105541 AU105541
27	54.2	50	10	10	AU105544 AU105544
28	54.2	50	10	10	AU105547 AU105547
29	54.2	50	10	10	AU105550 AU105550
30	54.2	50	10	10	AU105551 AU105551
31	54.2	50	10	10	AU105552 AU105552
32	54.2	50	10	10	AU105553 AU105553
33	54.2	50	10	10	AU105554 AU105554
34	54.2	50	10	10	AU105555 AU105555
35	54.2	50	10	10	AU105556 AU105556
36	54.2	50	10	10	AU105557 AU105557
37	54.2	50	10	10	AU105558 AU105558
38	54.2	50	10	10	AU105559 AU105559
39	54.2	50	10	10	AU105560 AU105560
40	54.2	50	10	10	AU105561 AU105561
41	54.2	50	10	10	AU105562 AU105562
42	54.2	50	10	10	AU105563 AU105563
43	54.2	50	10	10	AU105564 AU105564
44	54.2	50	10	10	AU105565 AU105565
45	54.2	50	10	10	AU105566 AU105566

#### ALIGNMENTS

RESULT 1  
A2793014 51 bp DNA GSS 16-FEB-2001  
LOCUS 2M0045E17R Mouse 10kb plasmid UGCG1M library Mus musculus genomic  
DEFINITION clone UGCG2M0045E17 R, DNA sequence.

ACCESSION A2793014.1 GI:12937364

VERSION A2793014.1

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A. and Wright,D., Weiss,R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

JOURNAL Unpublished (2000)

COMMENT Contact: Robert B. Weiss  
University of Utah Genome Center  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert length: 10000 Std Error: 0.00  
Plate: 0045 Row: E Column: 17  
Seq primer: CACACGAGAACAGTATGACC  
Class: plasmid ends  
High quality sequence stop: 51.  
Location/Qualifiers  
1. 51

FEATURES  
source

/organism="Mus musculus"  
/strain="C57BL/6j"  
/db\_xref="taxon:10090"  
/clone="UUCG2M0045E17"  
/clone\_lib="Mouse 10kb plasmid UUCG1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: PMD42nv; Purified genomic DNA from M. musculus C57BL/6j (male) was obtained from the Jackson Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g11473211419b1AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 15 a 12 c 13 g 11 t  
ORIGIN

Query Match 65.8%; Score 15.8; DB 13; Length 51;  
Best Local Similarity 89.5%; Pred. No. 5.9e+03;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtcacagcagacattc 19  
Db 27 gttccagcagcagcattat 45

RESULT 2  
AZ361084 59 bp DNA GSS 02-OCT-2000  
LOCUS 1M0104D08R Mouse 10kb plasmid UUCG1M library Mus musculus genomic  
DEFINITION Clone UUCG1M0104D08 R, DNA sequence.  
ACCESSION AZ361084  
VERSION AZ361084.1 GI:10474784  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 59)  
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamli,C.,  
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,  
'M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.  
and Wright,D., Weis,R.  
Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts  
Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: dunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0104 row: D column: 08  
Seq primer: CACACAGCAACAGCTATGACC  
Class: plasmid ends  
High quality sequence stop: 59.  
Location/Qualifiers

## FEATURES

## source

1. 59  
/organism="Mus musculus"  
/strain="C57BL/6j"  
/db\_xref="taxon:10090"  
/clone="UUCG1M0104D08"  
/clone\_lib="Mouse 10kb plasmid UUCG1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: PMD42nv; Purified genomic DNA from M. musculus C57BL/6j (male) was obtained from the Jackson Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g11473211419b1AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 15 a 9 c 18 g 17 t  
ORIGIN

Query Match 63.3%; Score 15.2; DB 13; Length 59;  
Best Local Similarity 85.0%; Pred. No. 1.1e+04;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 aagcagacgacattctgca 24  
Db 19 AAGCAGAGTCAGTCTCGCA 38

RESULT 3  
AZ921737 54 bp DNA GSS 20-MAR-2001  
LOCUS 100603JH12.y1 1006 - Rescuedu Grid G Zea mays genomic, DNA  
DEFINITION sequence.  
ACCESSION AZ921737  
VERSION AZ921737.1 GI:13393675  
KEYWORDS GSS.  
SOURCE Zea mays.  
ORGANISM Zea mays.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
clade; Panicoidae; Andropogoneae; Zea.  
1 (bases 1 to 54)  
Walbot,V.  
Maize genomic sequences found using engineered Rescuedu transposon  
Unpublished (2001)  
Contact: Walbot V  
Department of Biological Sciences  
Stanford University  
855 California Ave, Palo Alto, CA 94304, USA  
Tel: 650 723 2227  
Fax: 650 725 8221  
Email: walbot@stanford.edu  
Possible ligation site of ends cut by 2 different endonucleases.  
Reverse complemented post-ligation sequence from source sequence.  
Plate: 1006031 row: H column: 12  
Class: transposon-tagged.  
Location/Qualifiers

## FEATURES

1. 54  
/organism="Zea mays"  
/cultivar="mixed background W23/A188/B73"  
/db\_xref="taxon:4577"

/clone\_lib="1006 - Rescuemu Grid G"  
/tissue\_type="leaf"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Organ: leaf; Vector: Rescuemu (engineered from pBluescript backbone); Site\_1: BamHI; Site\_2: BglII; Rescuemu is a 4.9 kb, modified maize Mu transposon designed to allow plasmid rescue from total genomic DNA. Mu elements insert preferentially into transcription units. For more information on Rescuemu, go to the web site 'www.zmbl.iastate.edu' and follow the links for 'Rescuemu.' Grid G was grown at Stanford in 2000. DNA was extracted from leaf punches, double digested using BamHI and BglII, and ligated to form circular plasmids. DH10B cells were transformed and then screened on LB plates with ampicillin."

BASE COUNT 6 a 20 c 12 g 16 t

ORIGIN

Query Match 60.0%; Score 14.4; DB 13; Length 54;  
Best Local Similarity 75.0%; Pred. No. 2.4e+04;  
Matches 18; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 1 gcccaagcagagcaattctgca 24  
Db 38 GCGCCGAGGAGAGACTCTCTCCA 15

RESULT 4  
LOCUS BF339449 39 bp mRNA EST 22-NOV-2000  
DEFINITION 602039103F1 NCI\_CGAP\_Brn64 Homo sapiens CDNA clone IMAGE:4186752  
5', mRNA sequence.  
ACCESSION BF339449  
VERSION BF339449.1 GI:11285904  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 39)  
NIH-MGC http://mgc.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: David N. Louis, M.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLM9508 row: f column: 01  
High quality sequence stop: 38.  
Location/Qualifiers  
1. 39  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4186752"  
/clone\_lib="NCI\_CGAP\_Brn64"  
/tissue\_type="gliblastoma with EGFR amplification"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: Brain; Vector: PCMV-SPOK6; Site\_1: NotI; Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.57 kb. Constructed by Life Technologies. Note: this is a NCI\_CGAP Library."

BASE COUNT 6 a 11 c 13 g 9 t

ORIGIN

Query Match 58.3%; Score 14; DB 11; Length 39;  
Best Local Similarity 77.3%; Pred. No. 3.5e+04;  
Matches 17; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Best Local Similarity 77.3%; Pred. No. 3.5e+04;  
Matches 17; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 3 ccaagcagagcaattctgca 24  
Db 38 CCAAGCCTGAGCAGCTCTCCA 17

RESULT 5  
LOCUS BF342092 39 bp mRNA EST 22-NOV-2000  
DEFINITION 602012848F1 NCI\_CGAP\_Brn64 Homo sapiens CDNA clone IMAGE:4148962  
5', mRNA sequence.  
ACCESSION BF342092  
VERSION BF342092.1 GI:11288842  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 39)  
NIH-MGC http://mgc.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: David N. Louis, M.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLM9409 row: o column: 11  
High quality sequence stop: 37.  
Location/Qualifiers  
1. 39  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4148962"  
/clone\_lib="NCI\_CGAP\_Brn64"  
/tissue\_type="gliblastoma with EGFR amplification"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: brain; Vector: PCMV-SPOK6; Site\_1: NotI; Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.57 kb. Constructed by Life Technologies. Note: this is a NCI\_CGAP Library."

BASE COUNT 6 a 11 c 13 g 9 t

ORIGIN

Query Match 58.3%; Score 14; DB 11; Length 39;  
Best Local Similarity 77.3%; Pred. No. 3.5e+04;  
Matches 17; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 3 ccaagcagagcaattctgca 24  
Db 38 CCAAGCCTGAGCAGCTCTCCA 17

RESULT 6  
LOCUS A2658555 29 bp DNA GSS 14-DEC-2000  
DEFINITION IM0535113F Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
clone UUGC1M0535113 F. DNA sequence.  
ACCESSION A2658555  
VERSION A2658555.1 GI:11795617  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 29)



```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="U062M0181E10"
/clone_lib="Mouse 10kb plasmid U062M0181E10"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/notes="Vector: PMD42ny; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g11473211419b1AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT      8 a      17 c      5 g      13 t
ORIGIN

Query Match      55.8%; Score 13.4; DB 13; Length 43;
Best Local Similarity 93.3%; Pred No. 6.4e+04;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      10 agagcaattctgca 24
        ||||| |||||
Db      41 AGAGCAAGTTGCA 27

RESULT 9
LOCUS      AU013478      51 bp      mRNA      EST      03-AUG-1998
DEFINITION AU013478 Schizosaccharomyces pombe late log phase cDNA
ACCESSION  AU013478
VERSION     AU013478.1 GI:3368269
KEYWORDS    EST.
SOURCE      fission yeast.
ORGANISM    Schizosaccharomyces pombe
Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
Schizosaccharomycetaceae; Schizosaccharomycetaceae;
Schizosaccharomycetes.
1 (bases 1 to 51)
Moriyama, M. and Mita, K.
Identification of expressed sequence tags of Schizosaccharomyces
pombe
Unpublished (1998)
Contact: Mitsuo Moriyo
Genome Research Group
National Institute of Radiological Sciences
9-1, Anagawa-4-chome, Inage-ku, Chiba, Chiba 263-8555, Japan
Email: moriyo@nirs.go.jp
Location/Qualifiers
1. 51
/organism="Schizosaccharomyces pombe"
/db_xref="taxon:4896"
/clone="spc08212"
/clone_lib="Schizosaccharomyces pombe late log phase cDNA"
/sex="h minus"
/notes="Vector: M13mp19; The cDNA library of
Schizosaccharomyces pombe was prepared by cloning cDNA
into the SmaI site of M13mp19 DNA and the direction of DNA

```

```

sequences was not always from 5' to 3'. The cDNA data of
Schizosaccharomyces pombe are available for searching on
the World Wide Web. (URL: http://www.nirs.go.jp)

BASE COUNT      18 a      8 c      8 g      17 t
ORIGIN

Query Match      55.8%; Score 13.4; DB 10; Length 51;
Best Local Similarity 73.9%; Pred. No. 6.5e+04;
Matches 17; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy      2 tccaagcagagcaattctgca 24
        ||||| ||||| |||||
Db      14 TCCATGATGATGATTTTCGCA 36

RESULT 10
LOCUS      A2588457/c
DEFINITION A2588457 Mouse 10kb plasmid U062M0181E10 library Mus musculus genomic
ACCESSION  A2588457
VERSION     A2588457.1 GI:11710647
KEYWORDS    GSS.
SOURCE      house mouse.
ORGANISM    Mus musculus.
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 29)
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Relliy
and Wright, D., Weiss, R., Tinney, A., von Niederhausen, A.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: dunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0396 Row: L Column: 17
Seq primer: CACACAGAAACGATGACCC
Class: Plasmid ends
High quality sequence stop: 29.
Location/Qualifiers
1. 29
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="U062M0181E10"
/clone_lib="Mouse 10kb plasmid U062M0181E10"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/notes="Vector: PMD42ny; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g11473211419b1AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and

```

purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance.

BASE COUNT  
5 a 9 c 5 g 10 t

ORIGIN

Query Match 55.0%; Score 13.2; DB 13; Length 29;  
Best Local Similarity 83.3%; Pred. No. 7.7e+04;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 6 aggcagagcaattctgc 23  
||||| ||||| |||

Db 29 AGGAGGCAATTTCAGC 12

RESULT 11  
AZ632796/c

LOCUS AZ632796 42 bp DNA GSS 13-DEC-2000  
DEFINITION 1M0487E12R Mouse 10kb plasmid UGCG1M library Mus musculus genomic  
clone UGCG1M0487E12 R, DNA sequence.

ACCESSION AZ632796

VERSION AZ632796.1 GI:11754986

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 42)

AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A. and Wright,D., Weiss,R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb

JOURNAL Plasmid Inserts

COMMENT Unpublished (2000)

Contact: Robert B. Weiss  
University of Utah Genome Center  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: dunn@genetics.utah.edu  
Insert length: 10000 Std Error: 0.00  
Plate: 0487 row: E column: 12  
Seq primer: CACACAGCAACACGCTATGACC  
Class: Plasmid ends  
High quality sequence stop: 42.  
Location/Qualifiers  
1. 42  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UGCG1M0487E12"  
/clone\_lib="Mouse 10kb plasmid UGCG1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: PMD42uv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD42 (g1147321149b1aF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated

with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance.

BASE COUNT  
8 a 5 c 8 g 21 t

ORIGIN

Query Match 55.0%; Score 13.2; DB 13; Length 42;  
Best Local Similarity 83.3%; Pred. No. 7.8e+04;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 tccaagcagagcaattc 19  
||||| ||||| |||

Db 28 TCCAGGCAATTCCTGCA 11

RESULT 12

AL595919 57 bp mRNA EST 30-JUL-2001  
LOCUS AL595919 XGC-gastrula silurana tropicalis cdna clone Tcas004d19 5',  
DEFINITION mRNA sequence.

ACCESSION AL595919

VERSION AL595919.1 GI:15007994

KEYWORDS EST.

SOURCE western clawed frog.

ORGANISM Silurana tropicalis

REFERENCE 1 (bases 1 to 57)

AUTHORS Huckte,E., Taylor,R., McMurray,A., Ashurst,J.L., Zorn,A.M. and Rogers,J.

TITLE Sanger Xenopus tropicalis EST project 2001

JOURNAL Unpublished (2001)

COMMENT Contact: Huckie E

Sanger Centre  
Hinxton, Cambridgeshire, CB10 1SA, UK  
Email: tropesanger.ac.uk  
Sanger Xenopus tropicalis EST project 2001  
TROPICALIS\_SEQUENCE\_ID: Tgas004d19.sp6  
Sequencing primer: SP6  
This sequence is from a Xenopus Gene Collection (XGC) library constructed by Aaron M. Zorn.  
Location/Qualifiers  
1. 57  
/organism="Silurana tropicalis"  
/db\_xref="taxon:8364"  
/clone="Tgas004d19"  
/clone\_lib="XGC-gastrula"  
/dev\_stage="gastrula (stages 10-5-13 mixed)"  
/lab\_host="Escherichia coli DH10B"  
/note="Vector: pCS107; Site\_1: EcoRI; Site\_2: NotI; cDNA was oligo dt primed from 5ug of poly A+ RNA from stages 10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end."  
Location/Qualifiers

FEATURES  
source

1. 57  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UGCG1M0487E12"  
/clone\_lib="Mouse 10kb plasmid UGCG1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: PMD42uv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD42 (g1147321149b1aF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated

BASE COUNT 11 a 28 c 10 g 8 t

ORIGIN

Query Match 55.0%; Score 13.2; DB 10; Length 57;  
Best Local Similarity 83.3%; Pred. No. 7.9e+04;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 7 ggcagagcaattctgca 24  
||||| ||||| |||

Db 56 GGCAGGCAATTCCTGCA 39

RESULT 13

AU105510

LOCUS	AU105510	50 bp	mRNA		EST	05-APR-2001
DEFINITION	AU105510	Sugano Homo sapiens	CDNA library	Homo sapiens	CDNA clone	
ACCESSION	CAS08101	mRNA sequence.				
VERSION	AU105510					
KEYWORDS	AU105510.1	GI:13555031				
SOURCE	EST.					
ORGANISM	human.					
	Homo sapiens					
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.					
REFERENCE	1 (bases 1 to 50)					
AUTHORS	Suzuki,Y., Tsunoda,T., Taira,H., Mizushima-Sugano,J., Sese,J., Hata					
	,H., Ota,T., Isogai,T., Tanaka,T., Nakamura,Y., Morishita,S., Okubo					
	,K., Suyama,A. and Sugano,S.					
	Fine structural analysis of transcription start sites of human					
	mRNAs using full-length enriched and 5'-end enriched cDNA libraries					
	Unpublished (2001)					
	Contact: Yutaka Suzuki					
	Department of Virology					
	Institute of Medical Science, University of Tokyo					
	4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan					
	Email: yusuzuki@ims.u-tokyo.ac.jp					
	Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano					
	,S. Construction and characterization of a full length-enriched and					
	a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).					
FEATURES						
source	1..50					
	/organism="Homo sapiens"					
	/db_xref="taxon:9606"					
	/clone="CAS08101"					
	/clone_lib="Sugano Homo sapiens CDNA library"					
BASE COUNT	7 a	14 c	18 g	11 t		
ORIGIN						
	Query Match	54.2%;	Score 13;	DB 10;	Length 50;	
	Best Local Similarity	76.2%;	Pred. No. 9.6e+04;			
Matches	16;	Conservative	0;	Mismatches	5;	Indels 0; Gaps 0;
QY	2	tccaagcagacgaattctg	22			
	111	111111111111111				
Db	10	TCCGAGCGGGGCTATTCTCG	30			
RESULT 14						
LOCUS	AU105518	50 bp	mRNA		EST	05-APR-2001
DEFINITION	AU105518	Sugano Homo sapiens	CDNA library	Homo sapiens	CDNA clone	
ACCESSION	HRC01211	mRNA sequence.				
VERSION	AU105518					
KEYWORDS	AU105518.1	GI:13555039				
SOURCE	EST.					
ORGANISM	human.					
	Homo sapiens					
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.					
REFERENCE	1 (bases 1 to 50)					
AUTHORS	Suzuki,Y., Tsunoda,T., Taira,H., Mizushima-Sugano,J., Sese,J., Hata					
	,H., Ota,T., Isogai,T., Tanaka,T., Nakamura,Y., Morishita,S., Okubo					
	,K., Suyama,A. and Sugano,S.					
	Fine structural analysis of transcription start sites of human					
	mRNAs using full-length enriched and 5'-end enriched cDNA libraries					
	Unpublished (2001)					
	Contact: Yutaka Suzuki					
	Department of Virology					
	Institute of Medical Science, University of Tokyo					
	4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan					
	Email: yusuzuki@ims.u-tokyo.ac.jp					
	Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano					
	,S. Construction and characterization of a full length-enriched and					
	a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).					
FEATURES						
source	1..50					
	/organism="Homo sapiens"					
	/db_xref="taxon:9606"					
	/clone="CAS08101"					
	/clone_lib="Sugano Homo sapiens CDNA library"					
BASE COUNT	7 a	14 c	18 g	11 t		
ORIGIN						
	Query Match	54.2%;	Score 13;	DB 10;	Length 50;	
	Best Local Similarity	76.2%;	Pred. No. 9.6e+04;			
Matches	16;	Conservative	0;	Mismatches	5;	Indels 0; Gaps 0;
QY	2	tccaagcagacgaattctg	22			
	111	111111111111111				
Db	10	TCCGAGCGGGGCTATTCTCG	30			
RESULT 14						
LOCUS	AU105518	50 bp	mRNA		EST	05-APR-2001
DEFINITION	AU105518	Sugano Homo sapiens	CDNA library	Homo sapiens	CDNA clone	
ACCESSION	HRC01211	mRNA sequence.				
VERSION	AU105518					
KEYWORDS	AU105518.1	GI:13555039				
SOURCE						

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HRC01211"
/clone_lib="Sugano Homo sapiens cDNA library"
BASE COUNT      5 a      16 c      16 g      13 t
ORIGIN

Query Match      54.2%  Score 13;  DB 10;  Length 50;
Best Local Similarity 76.2%  Pred. No. 9.6e+04;
Matches 16;  Conservative 0;  Mismatches 5;  Indels 0;  Gaps 0;

QY      2  tccaagcagagcaattctctg  22
      ||| ||| | | ||| |||
Db      16  TCCGAGCGCGCCCTATTCTG  36

RESULT 15
AUI05525      50 bp  mRNA      EST      05-APR-2001
LOCUS
DEFINITION  AUI05525 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone
ACCESSION  HRC03805, mRNA sequence.
VERSION  AUI05525
KEYWORDS  AUI05525.1 GI:13555046
SOURCE  EST.
ORGANISM  human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 50)
Suzuki,Y., Tsunoda,T., Taira,H., Mizushima-Sugano,J., Sese,T., Hata
,H., Ota,T., Isogai,T., Tanaka,T., Nakamura,T., Morishita,S., Okubo
,K., Suyama,A. and Sugano,S.
Fine structural analysis of transcription start sites of human
mRNAs using full-length enriched and 5'-end enriched cDNA libraries
unpublished (2001)
Contact: Yutaka Suzuki
Department of Virology
Institute of Medical Science, University of Tokyo
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
Email: yusuzuki@ims.u-tokyo.ac.jp
Suzuki,Y., Yoshimoto-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano
,S. Construction and characterization of a full length-enriched and
a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).

FEATURES
Source
1..50
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HRC03805"
/clone_lib="Sugano Homo sapiens cDNA library"
BASE COUNT      6 a      15 c      15 g      14 t
ORIGIN

Query Match      54.2%  Score 13;  DB 10;  Length 50;
Best Local Similarity 76.2%  Pred. No. 9.6e+04;
Matches 16;  Conservative 0;  Mismatches 5;  Indels 0;  Gaps 0;

QY      2  tccaagcagagcaattctctg  22
      ||| ||| | | ||| |||
Db      15  TCCGAGCGCGCCCTATTCTG  35

Search completed: March 9, 2002, 00:09:28
Job time: 11044 sec

```

**THIS PAGE BLANK (USPTO)**

---